Exploring the Impact of Generative AI on Uncovering Drug-Related Issues in Individuals with Pre-existing Conditions

Anusha Guruprasad | Krystyn Gutu
Seidenberg School of CSIS, Pace University

Abstract

The project utilizes an innovative approach by generating synthetic data to gain insights into the factors underlying reactions in individuals with various pre-existing genetic conditions. This enables a more nuanced understanding of the diverse population affected by drug-related issues. Through the analysis of individual patient gene data, encompassing specific genetic markers, sophisticated Generative AI models are utilized to tailor drug treatments based on each patient's unique characteristics. This personalized medicine approach aims to optimize treatment outcomes while minimizing the risk of adverse drug reactions. The meticulous analysis and interpretation of the synthesized findings contribute to the advancement of personalized medicine approaches and underscore the critical role of advanced computational techniques, particularly Generative AI, in enhancing patient safety and driving innovation in clinical practice.

Related Work

Ietswaart et al. conducted a study on 2134 marketed drugs to analyze their secondary pharmacology and association with Adverse Drug Reactions (ADRs). Using FDA Adverse Event Reports, they developed random forest models to predict ADR occurrences from in vitro pharmacological profiles. They identified 221 target-ADR associations, indicating significant progress in understanding drug safety. This underscores the importance of computational techniques in enhancing patient safety in clinical practice.

Dataset

The ADReCS-Target dataset is a valuable resource comprising information on adverse drug reactions (ADRs) categorized by their target profiles. This dataset enables researchers and healthcare professionals to delve into the intricate mechanisms underlying ADRs, predict potential side effects of medications, and ultimately enhance drug safety and efficacy. Leveraging this dataset, I integrated open-source data focusing on ADRs associated with naturally occurring genetic variations. This enriched dataset includes essential columns such as unique identifiers, ADR codes, drug names, and variations, providing a comprehensive foundation for robust research models and deeper insights into the interplay between genetic factors and drug reactions**.**

	Unnamed: 0	BADD_TID	ADR_ID	ADReCS ID	ADR Term	Variation	Drug_Name
0	0	BADD_T04549	BADD_A03699	23.03.13.003	Rash macular	HLA-DRB1*16:02	co-trimoxazole
1	1	BADD_T04548	BADD_A03699	23.03.13.003	Rash macular	HLA-DRB1*16:01	co-trimoxazole
2	2	BADD_T04547	BADD_A03699	23.03.13.003	Rash macular	HLA-DRB1*15:01	EGF receptor inhibitors
3	3	BADD_T04546	BADD_A03699	23.03.13.003	Rash macular	HLA-DRB1*15:01	co-trimoxazole
4	4	BADD_T04545	BADD_A03699	23.03.13.003	Rash macular	HLA-DRB1*15	co-trimoxazole

Data Manipulation

The dataset, initially in XLSX format, was converted to CSV for easier processing. Preprocessing and manipulation were performed using Python. Some ADR_ID values were null or NAN, requiring their removal to ensure data integrity, given the difficulty in accurately imputing identification numbers.

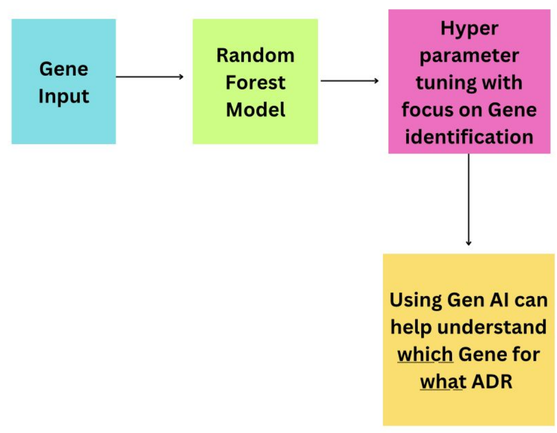
Methodology

Our analysis highlights correlations between drug names and adverse drug reaction (ADR) terms, suggesting shared reactions across medications, possibly due to undiagnosed genetic markers. Notably, drugs like Lumiracoxib and Lapatinib are linked to liver injury, indicating potential genetic vulnerabilities. Identifying genes associated with multiple ADRs provides insights into genetic susceptibilities.

Further investigation reveals a 100% correlation between liver injury and hyperphosphatasemia, suggesting a strong association worth exploring for its implications on patient care and drug safety.Our findings emphasize the importance of considering genetic factors in predicting and managing adverse drug reactions, informing personalized healthcare interventions and drug prescription practices. Healthcare professionals can tailor interventions to reduce risks and improve treatment plans, potentially saving lives, particularly for individuals lacking regular medical monitoring.

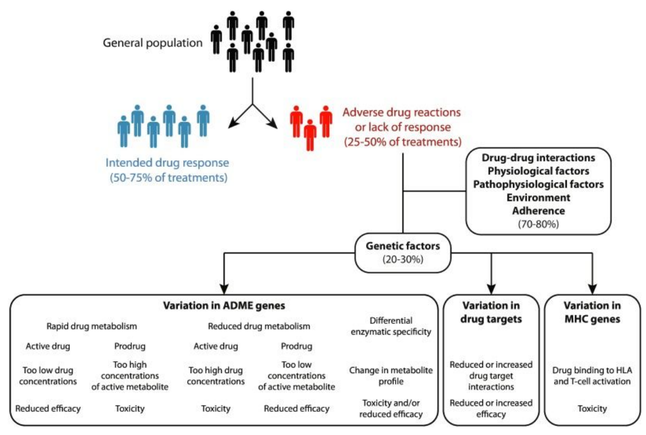
User Flow

The Random Forest algorithm, known for its versatility, is utilized in predictive modeling. It undergoes training on a provided dataset and generates predictions on a separate testing dataset. Accuracy is evaluated using the 'accuracy_score' function, offering a quantitative measure of performance. Hyperparameter tuning is crucial for optimizing Random Forest models. Grid Search Cross-Validation from scikit-learn is employed to explore different hyperparameter combinations. The best model is selected based on accuracy metrics, ensuring optimal performance. The integration of Gen AI (GPT 3.5) enhances predictive capabilities by identifying correlations between genetic variations and adverse drug reactions. This empowers healthcare professionals to tailor treatment plans and mitigate risks, ultimately improving patient outcomes and safety.



Conclusion & Use Cases

The hyperparameter tuning of the significantly boosted the Random Forest model's predictive capabilities, achieving over 80% accuracy and recall metrics. This enhanced analysis explores genetic variations' correlations with adverse drug reactions, enabling precise predictions based on individual genetic makeup using GenAI. Healthcare professionals can tailor treatment plans and mitigate risks, enhancing patient outcomes and safety through personalized medicine practices. This highlights the transformative potential of integrating Gen AI into healthcare predictive analytics.



References

- *Visan AI, Negut I. Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery. Life (Basel). 2024 Feb 7;14(2):233. doi: 10.3390/life14020233. PMID: 38398742; PMCID PMC10890405..*
- *Zhang P, Wang F, Hu J, Sorrentino R. Exploring the relationship between drug side-effects and therapeutic indications. AMIA Annu Symp Proc. 2013 Nov 16;2013:1568-77. PMID: 24551427; PMCID: PMC3900166.*
- *Ietswaart R, Arat S, Chen AX, Farahmand S, Kim B, DuMouchel W, Armstrong D, Fekete A, Sutherland JJ, Urban L. Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology. EBioMedicine. 2020 Jul;57:102837. doi: 10.1016/j.ebiom.2020.102837. Epub 2020 Jun 18. PMID:8 32565027; PMCID: PMC7379147.*
- *Biau, G., Scornet, E. A random forest guided tour. TEST 25, 197–227 (2016). https://doi.org/10.1007/s11749-016-0481-7*